Shuang Ni

+1 (514)549-0276 | ni_shuang@yahoo.com

SUMMARY

PhD candidate specializing in machine learning, dimensionality reduction, and applications to biological datasets. Skilled in uncovering insight from large-scale high-dimensional data, with a strong emphasis on responsible AI.
Seeking internship or part-time opportunities in machine learning research and engineering, data analysis, computational biology research, and quantitative analysis to leverage my expertise and drive impactful and reliable solutions.

EDUCATION

Université de Montréal, Montreal, Canada *09/2021 - 06/2026*
- PhD. in Computer Science (GPA: 4.30 /4.30) *(Anticipated)*
- Doctoral Researcher at Mila - Quebec AI Institute

Concordia University, Montreal, Canada *01/2019 - 06/2021*
- M.Sc. in Electrical and Computer Engineering (GPA: 4.30 /4.30)
- Transferred credits from the University of Manitoba (09/2017 - 12/2018):
  Statistical Aspect of Machine Learning, Parallel Processing, Advanced Signal Processing, Digital Image Processing

University of Electronic Science and Technology of China, Chengdu, China *09/2013 - 07/2017*
- Bachelor of Electronic Information Engineering (GPA: 3.47/4.00)

SKILLS

- **Programming:** Python, R, PyTorch, TensorFlow, Scikit-learn, C, C++.
- **Tools**: MATLAB, Mathematica, Linux, GitHub.

EXPERIENCE

Doctoral Research Associate | Mila - Quebec AI Institute - Montreal, Canada

Capture Continuous Population Structure in Human Genomic Biobanks *05/2024 - present*
- Designed dimensionality reduction methods combining autoencoders and PHATE for multiple human genetic biobanks.
- Applied archetypal analysis to estimate admixture signals in populations, highlighting genetic contributions from ancestral sources.
- Performed quantitative comparisons with other dimensionality reduction techniques, showcasing improved performance of the proposed method in capturing continuous genetic variations and clustering populations based on genetic similarity.

Modeling Dynamic Mechanisms of Drug Treatment Effects with GeoSinkhorn Flow *05/2024 - present*
- Collaborated on the design and validation of the GeoSinkhorn Flow model, assessing its capability to model geodesic distances and dynamic transport across distributions.
- Conducted experiments on synthetic manifolds to demonstrate the model's ability to simulate geodesic distances on manifolds.
- Evaluated the model on real-world datasets, predicting drug combination effects and temporal treatment responses, demonstrating its effectiveness in dynamic modeling.

Enhancing Semi-Supervised Visualization with Autoencoders and Random Forest Proximities *09/2023 - present*
- Designed and implemented various novel architectures combining geometry-regularized autoencoders with Random Forest-based PHATE to extend embeddings for out-of-sample data points while preserving the intrinsic manifold structure.
- Introduced a proximity-based landmark selection method, reducing training time by 40% without compromising extension quality.
- Conducted experiments on large-scale datasets, achieving consistent embedding quality using only 10% of the training data.

Machine Learning-Based Biomarker Discovery and Cellular Landscape Analysis in Multiple Sclerosis (MS) *09/2022 - present*
- Corrected batch effects and integrated scRNA-seq data from multiple patients of MS and other related diseases.
- Applied PHATE and Multiscale PHATE to visualize and analyze cell types and subtypes.
- Developed predictive models using clinical and immunological data to identify biomarkers and disease progression indicators.

Graduate Research Assistant | Concordia University - Montreal, Canada

Deep Learning Mechanism of Revenue Maximization in Mobile Edge Computing *09/2019 - 04/2021*
- Designed and implemented a revenue maximization incentive mechanism using deep learning techniques.
- Developed a custom neural network model with TensorFlow to optimize revenue in the context of cooperative task offloading within a mobile edge computing system.
- Formulated a Lagrangian function as a loss function to effectively address the revenue maximization problem while considering three practical constraints.

Muscle Oxygen Saturation Quantitative Measurement and Error Detection *09/2017 - 09/2020*

- Developed a methodology to non-invasively and quantitatively measure muscle oxygen saturation (SmO2) using 5-wavelength diffuse reflectance continuous-wave near-infrared spectroscopy (NIRS).
- Employed non-linear least squares fitting of the 5-wavelength measured attenuation spectrum to a Taylor expansion attenuation model, enabling the determination of SmO2 values for calf muscles in various workout statuses.
- Utilized SVM with an RBF kernel in Scikit-learn to classify labeled attenuation spectrum datasets and predict the accuracy of collected data.
- Designed an application for SmO2 calculation and fault diagnostics using MATLAB App Designer.

PUBLICATIONS

- S. Ni, A. Aumon, G. Wolf, K. R. Moon and J. S. Rhodes, "Enhancing Supervised Visualization Through Autoencoder and Random Forest Proximities for Out-of-Sample Extension," *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*, London, United Kingdom, 2024, pp. 1-6
- G. Li, J. Cai and S. Ni, "Truthful Deep Mechanism Design for Revenue-Maximization in Edge Computing with Budget Constraints," in *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 902-914, Jan. 2022

HONORS & AWARDS

- Bourse d'exemption UdeM, 2021
- Concordia Merit Scholarship, 2019
- International Graduate Student Entrance Scholarship of University of Manitoba, 2017
- Third-class Scholarship of UESTC, 2016
- Outstanding Volunteer of UESTC, 2014